

Zephyr: Direct Distillation of LM Alignment

Year: 2023 | Citations: 507 | Authors: Lewis Tunstall, Edward Beeching, Nathan Lambert, Nazneen Rajani, Kashif Rasul

Abstract

We aim to produce a smaller language model that is aligned to user intent. Previous research has shown that applying distilled supervised fine-tuning (dSFT) on larger models significantly improves task accuracy; however, these models are unaligned, i.e. they do not respond well to natural prompts. To distill this property, we experiment with the use of preference data from AI Feedback (AIF). Starting from a dataset of outputs ranked by a teacher model, we apply distilled direct preference optimization (dDPO) to learn a chat model with significantly improved intent alignment. The approach requires only a few hours of training without any additional sampling during fine-tuning. The final result, Zephyr-7B, sets the state-of-the-art on chat benchmarks for 7B parameter models, and requires no human annotation. In particular, results on MT-Bench show that Zephyr-7B surpasses Llama2-Chat-70B, the best open-access RLHF-based model. Code, models, data, and tutorials for the system are available at <https://github.com/huggingface/alignment-handbook>.